

## Schriftliche Prüfung (90 Minuten)

### Bemerkungen:

- Alle schriftlichen Hilfsmittel und ein Taschenrechner sind erlaubt.
- Mobiltelefone sind auszuschalten!
- Lesen Sie zuerst alle Aufgaben durch! Verweilen Sie nicht zu lange bei einem Aufgabenteil, der Ihnen grosse Schwierigkeiten bereitet! Es wird nicht erwartet, dass Sie alle Aufgaben vollständig lösen.
- Wenn nicht anders vermerkt, sind die Tests auf dem 5%-Niveau durchzuführen.
- Der Lösungsweg muss (ausser bei den Multiple-Choice-Aufgaben) immer ersichtlich sein.
- Bei den Multiple-Choice-Aufgaben ist jeweils genau eine Antwort korrekt. Eine korrekte Antwort gibt 1 **Pluspunkt** und eine falsche Antwort  $\frac{1}{2}$  **Minuspunkt**. Minimal erhält man für eine ganze Multiple-Choice Aufgabe 0 Punkte. Tragen Sie die korrekten Antworten der Multiple-Choice-Aufgaben mit Kreuzchen in das zugehörige Antwortblatt ein.
- Die nötigen Tabellen befinden sich auf den hintersten Seiten dieser Prüfung.

**Viel Erfolg!**

1. (11 Punkte) P. Baader, der Besitzer einer Badeanstalt am Zürichsee, verzeichnet dieses Jahr deutlich weniger Badegäste als im Vorjahr. Er vermutet als Hauptursache, dass die Seetemperatur im Jahr 2012 deutlich höher war als im aktuellen Jahr 2013. Falls er statistisch korrekt nachweisen kann, dass die durchschnittliche monatliche Temperatur im Vorjahr mehr als  $2^\circ\text{C}$  höher war als in diesem Jahr, kann er bei der Freibadvereinigung Anspruch auf eine Schlechtwetter-Entschädigung geltend machen.

Er beauftragt Sie, einen geeigneten statistischen Test durchzuführen. Zu diesem Zweck stellt er Ihnen die folgende Tabelle mit den durchschnittlichen Seetemperaturen für die Monate März bis Juli der Jahre 2012 und 2013 zur Verfügung:

	März	April	Mai	Juni	Juli
Jahr 2012	7.1	11.8	15.1	19.7	21.0
Jahr 2013	5.6	9.3	12.2	16.5	18.6
Differenz	1.5	2.5	2.9	3.2	2.4

Nehmen Sie an, dass die monatlichen Differenzen Seetemperatur im Jahr 2012 minus Seetemperatur im Jahr 2013 unabhängig voneinander normalverteilt mit Erwartungswert  $\mu_D$  und Standardabweichung  $\sigma_D$  sind. Aus den Daten wurden  $\hat{\mu}_D = \bar{D} = 2.5$  und  $\hat{\sigma}_D = s_D = 0.64$  geschätzt.

- Es handelt sich um einen gepaarten Test. Warum?
- Geben Sie die Null- und die Alternativhypothese an und begründen Sie kurz Ihre Wahl.
- Führen Sie den geeigneten  $t$ -Test zum Niveau 0.05 durch: Bestimmen Sie den Wert der Teststatistik  $T$ , die Verteilung der Teststatistik  $T$  unter der Nullhypothese, den Verwerfungsbereich für  $T$  und den Testentscheid.

Nehmen Sie in den folgenden Teilaufgaben d) und e) an, die Standardabweichung der Differenz  $\sigma_D = 0.64$  sei bekannt und muss *nicht* aus den Daten geschätzt werden.

- Wie ist die Teststatistik des  $z$ -Tests unter der Nullhypothese verteilt? Geben Sie Verteilungsfamilie und Parameter an.
- Bestimmen Sie ein einseitiges (passend zur Wahl der Alternativhypothese) 95%-Vertrauensintervall für  $\mu_D$ . Verwirft der  $z$ -Test die Nullhypothese auf dem 5%-Niveau? Begründen Sie kurz.

P. Baader ist über das Resultat des  $t$ -Tests nicht erfreut. Er überlegt sich deshalb, jeden gemessenen Wert im Jahr 2013 genau um soviel zu verkleinern, dass sich der Testentscheid des  $t$ -Tests zu seinen Gunsten ändert. Sie erhalten von ihm die folgenden Daten:

	März	April	Mai	Juni	Juli
Jahr 2012	7.1	11.8	15.1	19.7	21.0
Jahr 2013	$5.6 - c$	$9.3 - c$	$12.2 - c$	$16.5 - c$	$18.6 - c$
Differenz	$1.5 + c$	$2.5 + c$	$2.9 + c$	$3.2 + c$	$2.4 + c$

- Wie gross muss  $c$  mindestens sein, damit sich der Testentscheid des  $t$ -Tests ändert? (*Hinweis: Überlegen Sie sich, wie sich  $\bar{D}$  und  $\hat{\sigma}_D$  ändern, wenn Sie bei jeder monatlichen Differenz den Wert  $c$  addieren. Benutzen Sie danach die Teststatistik des  $t$ -Tests.*)

**2. (9 Punkte)** Johannes möchte im Rahmen seiner Masterarbeit eine Umfrage unter ETH-Studenten durchführen. Er geht davon aus, dass Studenten hilfsbereit sind und jeder Student die verschickten Fragebogen mit Wahrscheinlichkeit  $\frac{1}{2}$  beantwortet. Er schickt die Umfrage an 8 zufällig ausgewählte Studenten. Sei  $X \in \{0, 1, \dots, 8\}$  die Anzahl beantworteter Umfragen.

- a) Wie ist  $X$  verteilt? Geben Sie die Verteilungsfamilie und die Parameter an.
- b) Wie gross ist die Wahrscheinlichkeit, dass mindestens 7 Umfragen beantwortet werden?
- c) Wie gross ist die Wahrscheinlichkeit, dass genau ein Viertel der angeschriebenen Studenten die Umfrage beantwortet? Wie gross ist die Wahrscheinlichkeit, dass genau ein Viertel der Studenten die Umfrage *nicht* beantwortet?

Von den 8 angeschriebenen Studenten haben nur 2 die Umfrage ausgefüllt und zurückgeschickt. Johannes befürchtet nun, dass seine Annahme, dass die Studenten die Umfrage mit Wahrscheinlichkeit  $\frac{1}{2}$  beantworten, zu optimistisch war und sie in Wirklichkeit kleiner ist. Er möchte dies anhand eines exakten statistischen Tests (auf dem Signifikanzniveau 0.05) überprüfen.

- d) Was sind die Null- und die Alternativhypothese?
- e) Führen Sie einen geeigneten exakten Test durch. Geben Sie den Verwerfungsbereich für  $X$  und den Testentscheid an.
- f) Berechnen Sie die Wahrscheinlichkeit für einen Fehler 2. Art, falls der Parameter der Binomialverteilung in Tat und Wahrheit  $\pi = 0.25 \in H_A$  ist. Wie müsste das Signifikanzniveau  $\alpha$  verändert werden, damit der Fehler 2. Art kleiner wird?

Nach diesem Test beschliesst Johannes, seine Umfrage an 500 Studenten zu schicken. Er erhält von 211 Studenten eine Antwort.

- g) Geben Sie ein approximatives zweiseitiges 95%-Vertrauensintervall für den unbekannt Parameter  $\pi$  an.

3. (7 Punkte) Oskar und Max werfen nacheinander einen Ball von unterschiedlich hohen Positionen (hoehe) auf den Boden und messen die Zeit (zeit) bis der Ball aufschlägt. Folgendes Modell wurde an die Daten angepasst:

$$\text{hoehe}_i = \beta_0 + \beta_1 \cdot \text{zeit}_i + \beta_2 \cdot \text{zeit}_i^2 + \varepsilon_i, \quad \varepsilon_i \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2).$$

Der (unvollständige) Regressionsoutput sieht wie folgt aus:

Residuals:

Min	1Q	Median	3Q	Max
-2.359	-0.692	-0.185	0.926	2.522

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1.987	2.572	???	???
zeit	???	2.326	-0.82	0.42
zeit.quadrat	5.340	0.492	10.86	2.9e-12

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

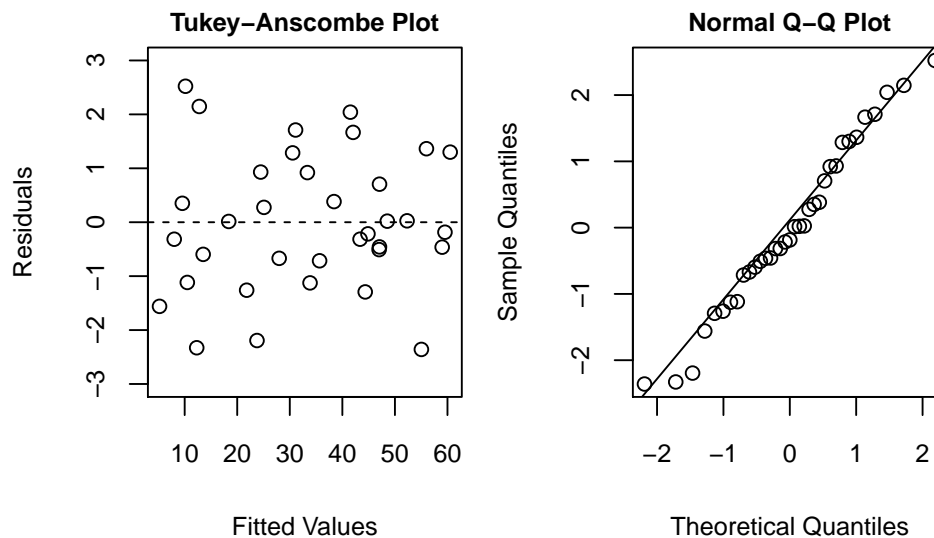
Residual standard error: 1.31 on 32 degrees of freedom

Multiple R-squared: 0.994, Adjusted R-squared: 0.994

F-statistic: 2.81e+03 on 2 and 32 DF, p-value: <2e-16

- 1) Wie viele Messungen haben Oskar und Max gemacht?
  - a) 30
  - b) 32
  - c) 34
  - d) 35
- 2) Wird die Nullhypothese  $H_0 : \beta_0 = 0$  auf dem 5%-Niveau verworfen?
  - a) Ja.
  - b) Nein.
  - c) Keine Angabe möglich.
- 3) Wie gross ist die Schätzung von  $\hat{\beta}_1$ ?
  - a) -2.84
  - b) -1.91
  - c) -0.35
  - d) -0.34
- 4) Berechnen Sie das exakte zweiseitige 99%-Konfidenzintervall für  $\beta_2$ .
  - a) [3.99, 6.69]
  - b) [4.34, 6.34]
  - c) [4.36, 6.32]
  - d) [4.59, 5.99]

- 5) Oskar und Max werfen den Ball aus 20 Metern Höhe herunter und messen dabei 2 Sekunden bis zum Aufprall. Wie gross ist der Fehler gemäss dem Modell ( $\hat{=}$  Residuum)?
- 1.4
  - 0.5
  - 1.4
  - 0.5
- 6) Handelt es sich bei dem obigen Modell um ein lineares Modell?
- Nein, weil das Modell einen quadratischen Regressionsterm (`zeit.quadrat`) enthält.
  - Nein, weil die Variable `zeit` zweifach als erklärende Variable im Modell vorkommt.
  - Ja, weil das Modell linear in den Koeffizienten ist.
  - Ja, weil die Zielvariable (`hoehe`) nicht transformiert ist.
- 7) Betrachten Sie die nachfolgenden Plots. Welche der folgenden Aussagen trifft zu?
- Alle Modellannahmen sind erfüllt.
  - Die Fehlervarianz ist nicht konstant, aber die Normalverteilungsannahme ist plausibel.
  - Die Fehlervarianz ist konstant, aber die Normalverteilungsannahme trifft nicht zu.
  - Sowohl konstante Fehlervarianz als auch Normalverteilungsannahme treffen nicht zu.



4. (9 Punkte) Die folgenden Aufgaben sind zufällig angeordnet und insbesondere nicht nach Schwierigkeitsgrad sortiert.

1) Seien  $X \sim \mathcal{N}(2, 16)$  und  $Y \sim \text{Poisson}(3)$  zwei unabhängige Zufallsvariablen. Berechnen Sie  $\text{Var}(X + 3Y - 5)$ .

- a) 20
- b) 25
- c) 30
- d) 38
- e) 43

2) Sei  $X \sim \text{Uniform}([-2; 2])$ , dann sind  $X$  und  $3 - X$

- a) Unabhängig und unkorreliert.
- b) Abhängig aber unkorreliert.
- c) Unabhängig aber korreliert.
- d) Abhängig und korreliert.

3) Sei  $X \sim \text{Exp}(1/3)$  eine Zufallsvariable. Berechnen Sie  $P[X \geq 5]$ .

- a) 0
- b) 0.19
- c) 0.76
- d) 0.81
- e) 0.94

4) Bei welcher der folgenden Funktionen handelt es sich um eine Dichtefunktion?

a)

$$f(x) = 0$$

b)

$$f(x) = \begin{cases} 0 & \text{für } x \leq 0 \\ 2x & \text{für } x \in (0; 1) \\ 1 & \text{für } x \geq 1 \end{cases}$$

c)

$$f(x) = \begin{cases} -4x^3 & \text{für } x \in [-1; 0] \\ 0 & \text{sonst} \end{cases}$$

d)

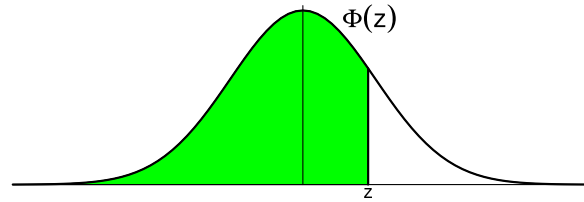
$$f(x) = \begin{cases} -1 & \text{für } x \in [0; 1] \\ 1 & \text{für } x \in (1; 3] \\ 0 & \text{sonst} \end{cases}$$

5) Aus einer Stichprobe von 2 Werten berechnet man den Mittelwert ( $\bar{x} = 5.5$ ) und die empirische Varianz ( $\hat{\sigma}^2 = 4.5$ ). Der kleinere der beiden Werte ist dann:

- a) -1
- b) 1
- c) 3.25
- d) 4
- e) Es gibt nicht genügend Informationen, um die Frage zu beantworten.

- 6) Ueli untersucht, ob das Geschlecht einer Person einen Einfluss auf die Orientierungsfähigkeit hat. Er testet 50 Frauen und 50 Männer und stellt dann die folgenden Hypothesen auf:  
 $H_0$ : Es gibt keinen Unterschied in der Orientierungsfähigkeit von Mann und Frau.  
 $H_A$ : Es gibt einen Unterschied in der Orientierungsfähigkeit von Mann und Frau.  
 Diese Hypothesen möchte Ueli nun anhand eines ungepaarten, 2-seitigen  $t$ -Tests auf dem 5%-Niveau analysieren. Der  $p$ -Wert des Testergebnisses beträgt 0.99. Ueli schliesst daraus, dass es mit hoher Wahrscheinlichkeit keinen Unterschied in der Orientierungsfähigkeit von Mann und Frau gibt. Welchen Fehler hat Ueli gemacht?
- Der Test hätte gepaart durchgeführt werden sollen.
  - Die Alternativhypothese hätte einseitig gewählt werden müssen.
  - Das Testergebnis ( $p$ -Wert= 99%) ist signifikant auf dem Niveau  $\alpha = 5\%$ , folglich hat das Geschlecht einen Einfluss.
  - Wenn die Nullhypothese akzeptiert wird, kann Ueli nicht mit hoher Wahrscheinlichkeit schliessen, dass sie richtig ist.
- 7) Sei  $X \sim \mathcal{N}(3, 8)$  eine Zufallsvariable. Berechnen Sie  $P[X = 5]$ .
- 0
  - 0.24
  - 0.76
  - 0.88
  - 0.999
- 8) Für zwei Ereignisse  $A$  und  $B$  hat man  $P(A) = 0.8$  und  $P(B) = 0.9$ . Welche Werte kann  $P(A \cap B)$  ausschliesslich annehmen?
- $0 \leq P(A \cap B) \leq 1$
  - $0.5 \leq P(A \cap B) \leq 0.9$
  - $0.7 \leq P(A \cap B) \leq 0.9$
  - $0.7 \leq P(A \cap B) \leq 0.8$
  - $0.8 \leq P(A \cap B) \leq 0.9$
- 9) Eine Stichprobe besteht aus  $n$  Beobachtungen. Der Mittelwert der Stichprobe beträgt 7. Die Beobachtung  $x_1 = 31$  wird jetzt gestrichen. Danach ist der Mittelwert der restlichen Beobachtungen  $x_2, \dots, x_n$  gleich 5. Wie viele Beobachtungen beinhaltet die ursprüngliche Stichprobe?
- $n = 5$
  - $n = 12$
  - $n = 13$
  - $n = 24$
  - $n = 57$

Tabelle der Kumulativen Normalverteilung  $\Phi(z) = P[Z \leq z]$ ,  $Z \sim \mathcal{N}(0, 1)$

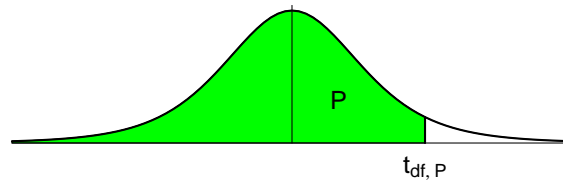


Bsp.:  $P[Z \leq 1.96] = 0.975$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998



## Perzentile der t-Verteilung



Bsp.:  $t_{9; 0.975} = 2.262$

$df$	$t_{0.60}$	$t_{0.70}$	$t_{0.80}$	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$	$t_{0.995}$
1	0.325	0.727	1.376	3.078	6.314	12.706	31.821	63.657
2	0.289	0.617	1.061	1.886	2.920	4.303	6.965	9.925
3	0.277	0.584	0.978	1.638	2.353	3.182	4.541	5.841
4	0.271	0.569	0.941	1.533	2.132	2.776	3.747	4.604
5	0.267	0.559	0.920	1.476	2.015	2.571	3.365	4.032
6	0.265	0.553	0.906	1.440	1.943	2.447	3.143	3.707
7	0.263	0.549	0.896	1.415	1.895	2.365	2.998	3.499
8	0.262	0.546	0.889	1.397	1.860	2.306	2.896	3.355
9	0.261	0.543	0.883	1.383	1.833	2.262	2.821	3.250
10	0.260	0.542	0.879	1.372	1.812	2.228	2.764	3.169
11	0.260	0.540	0.876	1.363	1.796	2.201	2.718	3.106
12	0.259	0.539	0.873	1.356	1.782	2.179	2.681	3.055
13	0.259	0.538	0.870	1.350	1.771	2.160	2.650	3.012
14	0.258	0.537	0.868	1.345	1.761	2.145	2.624	2.977
15	0.258	0.536	0.866	1.341	1.753	2.131	2.602	2.947
16	0.258	0.535	0.865	1.337	1.746	2.120	2.583	2.921
17	0.257	0.534	0.863	1.333	1.740	2.110	2.567	2.898
18	0.257	0.534	0.862	1.330	1.734	2.101	2.552	2.878
19	0.257	0.533	0.861	1.328	1.729	2.093	2.539	2.861
20	0.257	0.533	0.860	1.325	1.725	2.086	2.528	2.845
21	0.257	0.532	0.859	1.323	1.721	2.080	2.518	2.831
22	0.256	0.532	0.858	1.321	1.717	2.074	2.508	2.819
23	0.256	0.532	0.858	1.319	1.714	2.069	2.500	2.807
24	0.256	0.531	0.857	1.318	1.711	2.064	2.492	2.797
25	0.256	0.531	0.856	1.316	1.708	2.060	2.485	2.787
26	0.256	0.531	0.856	1.315	1.706	2.056	2.479	2.779
27	0.256	0.531	0.855	1.314	1.703	2.052	2.473	2.771
28	0.256	0.530	0.855	1.313	1.701	2.048	2.467	2.763
29	0.256	0.530	0.854	1.311	1.699	2.045	2.462	2.756
30	0.256	0.530	0.854	1.310	1.697	2.042	2.457	2.750
31	0.255	0.530	0.853	1.309	1.696	2.040	2.452	2.744
32	0.255	0.530	0.853	1.309	1.694	2.037	2.449	2.738
33	0.255	0.530	0.853	1.308	1.693	2.035	2.445	2.733
34	0.255	0.529	0.852	1.307	1.691	2.032	2.441	2.728
35	0.255	0.529	0.852	1.306	1.690	2.030	2.438	2.724
40	0.255	0.529	0.851	1.303	1.684	2.021	2.423	2.704
60	0.254	0.527	0.848	1.296	1.671	2.000	2.390	2.660
90	0.254	0.526	0.846	1.291	1.662	1.987	2.368	2.632
120	0.254	0.526	0.845	1.289	1.658	1.980	2.358	2.617
$\infty$	0.253	0.524	0.842	1.282	1.645	1.960	2.326	2.576